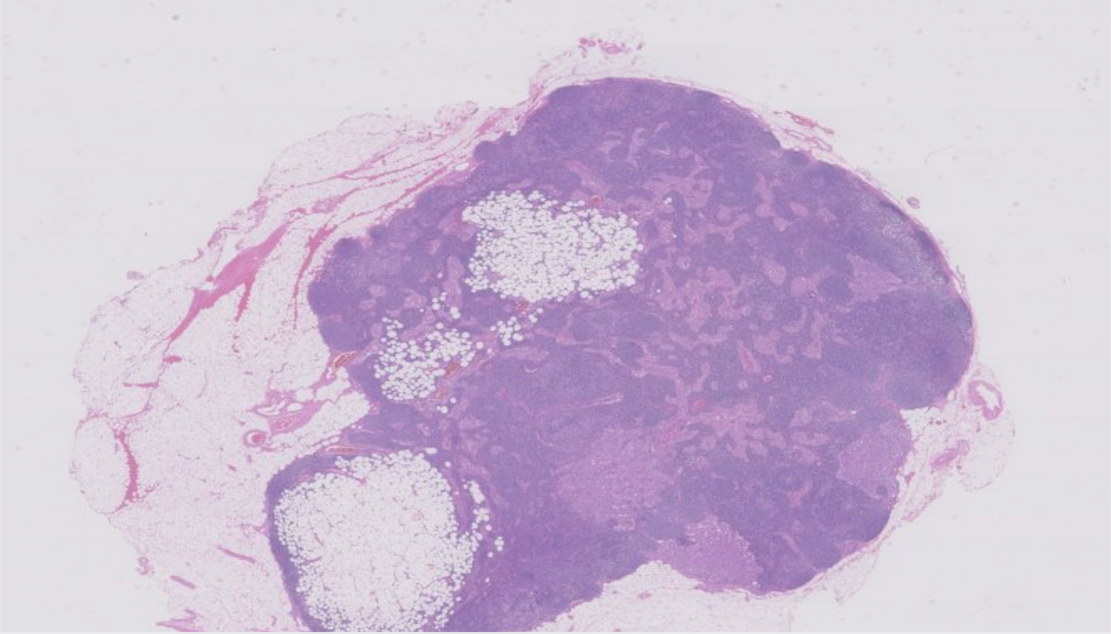


Detecting Cancer Metastases on Gigapixel Pathology Images

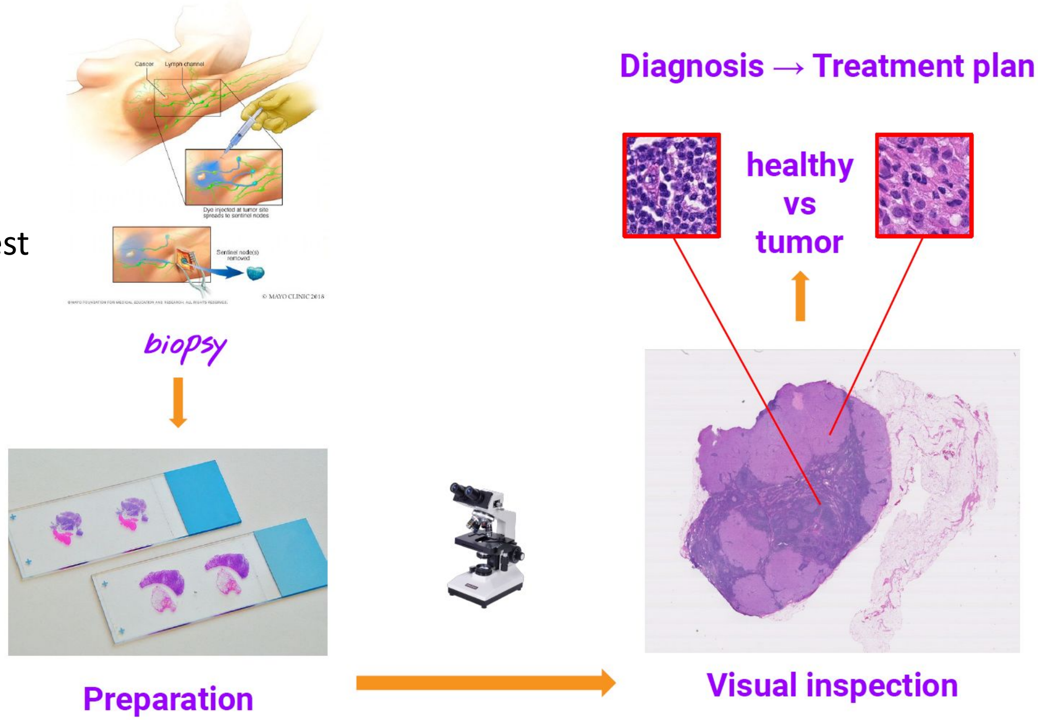
Team Members: Chenxi Di, Tianqi Lou

Background

Visual inspection

● Can be tedious and error prone (despite best efforts of pathologists).

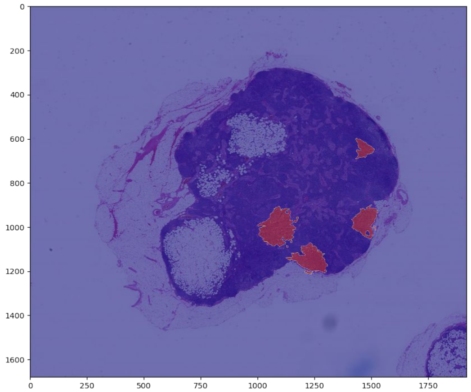
● Tissue samples are volumes: when sliced, many images / sample.



Goal

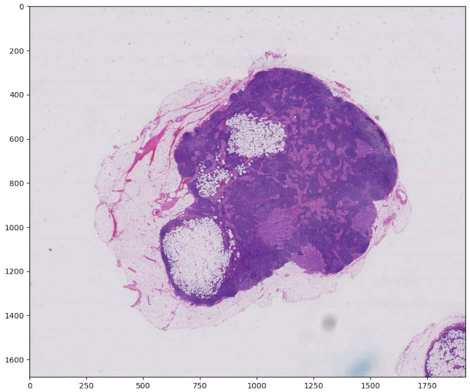
Given a collection of training data, develop a model that outputs a Heatmap showing regions of a biopsy image likely to contain cancer.

Assist, not automate:
insert into workflow as an automatic second opinion, with enough training data, we can develop models radically reduce misdiagnosis at little cost.



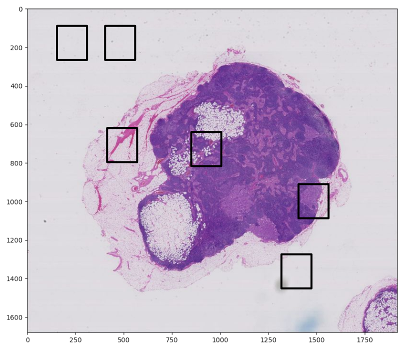
Ground Truth

(Red region: has tumor)

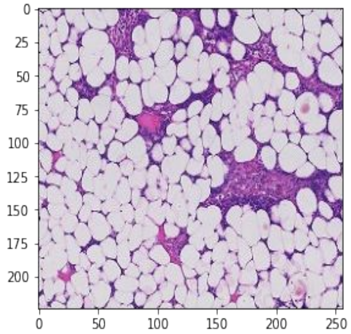


slide

Methodology



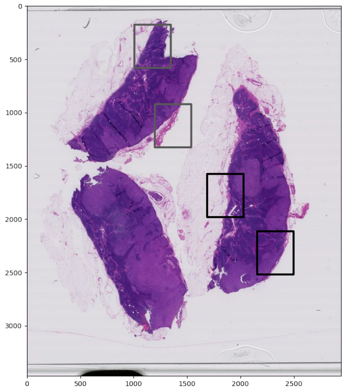
1. Patch extraction



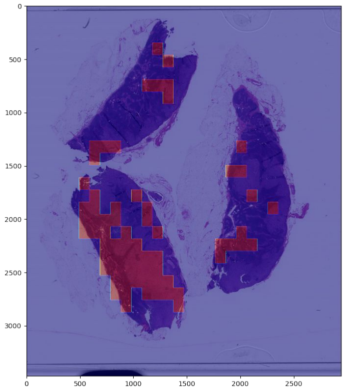
2. Train with deep learning model

a. Single scale fully connected model or

b. Multi scale fully connected model

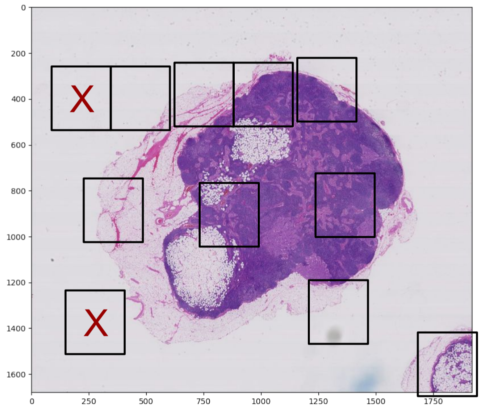


3. Use trained model to predict the probability of containing tumor cell in each sliding windows



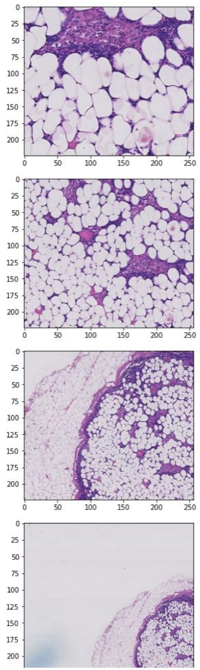
4. Based on the probability, reconstruct the predicted ‘heatmap’

Patch Extraction

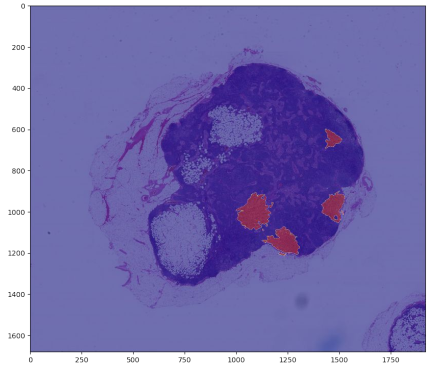


Sliding window over the entire slide

Check if the patch has >50% tissue



For each patch, extract more patches at different zoom levels(level = 3,4,5,6)



Compare with corresponding tumor mask file to check if each patch includes tumor cell.

Repeat this process and generate labeled data set for all 10 slides.

Model

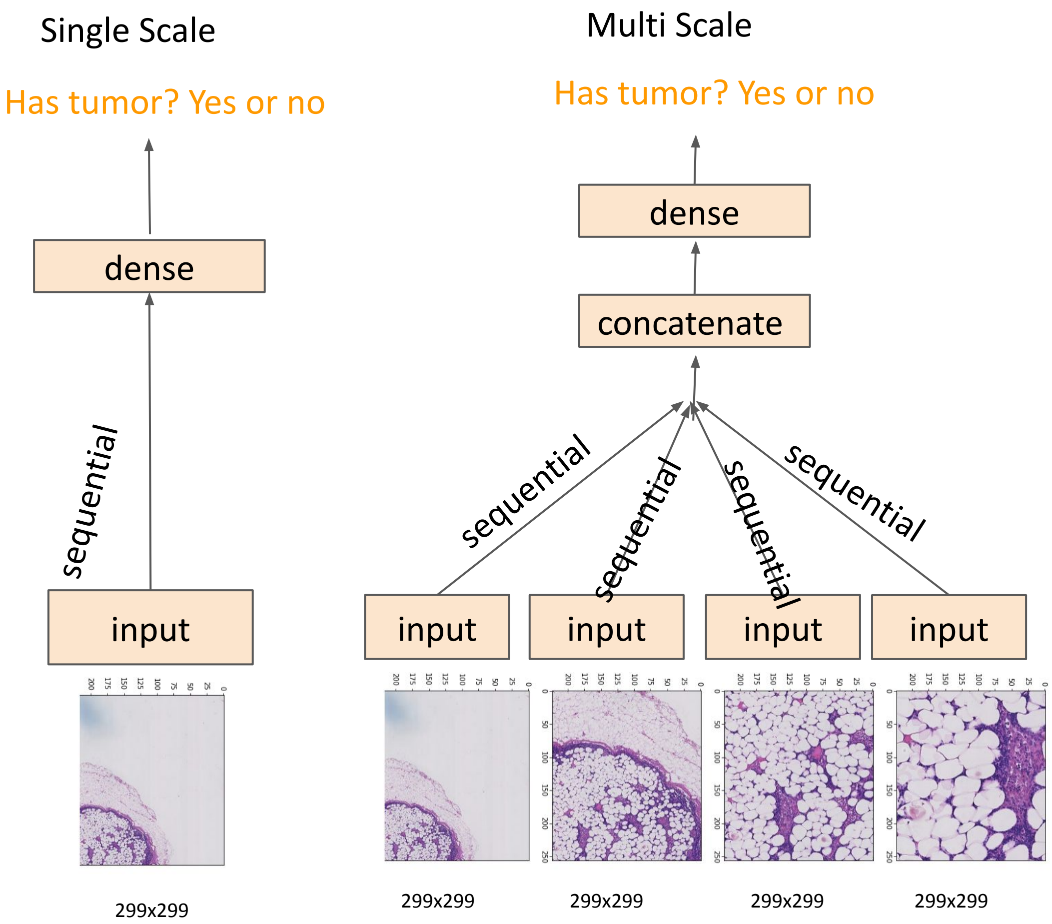
Inspired from paper, we implemented two models as below:

- Single Scale Model: InceptionV3
- Multi Scale Model: InceptionV3/VGG16 with 4 different zoom levels inputs.

Imbalanced data solution:

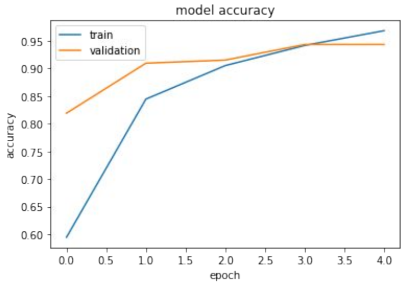
1) Randomly select healthy patches with probability of 0.5

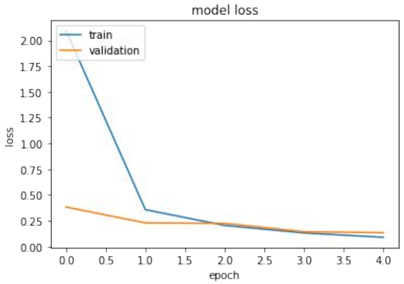
2) Use data augmentation inspired from the paper(color perturbations and flip/rotate augmentation) to increase the proportion of tumor patches.



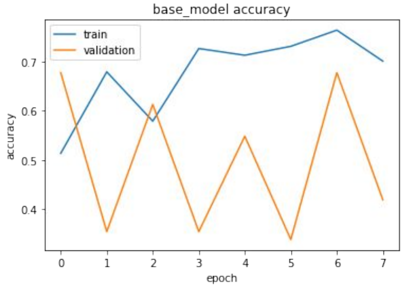
Result

Multi-scale model(InceptionV3): (test set accuracy: 0.96)

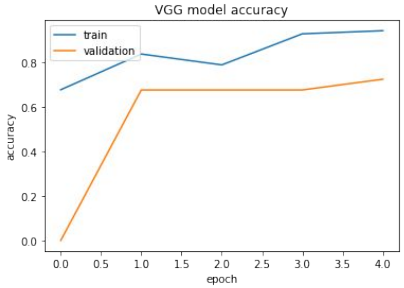


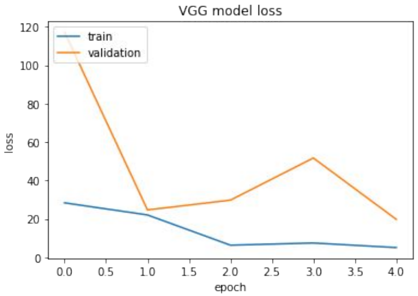


Single-scale model(InceptionV3):

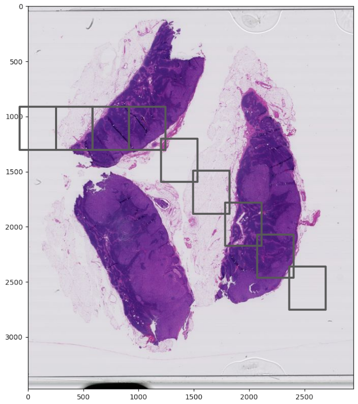


Multi-scale model(VGG): (test set accuracy 0.88)



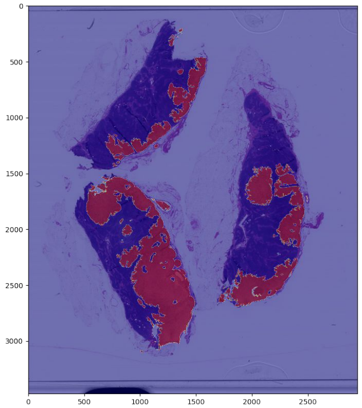


Heatmap Reconstruction

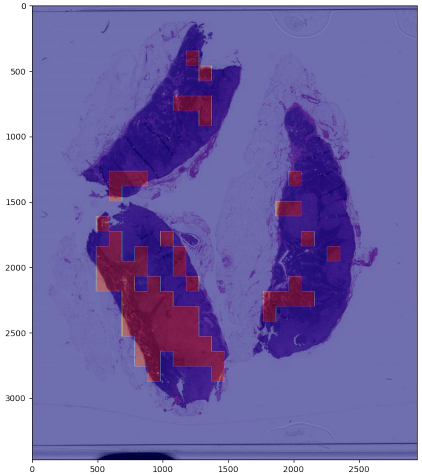


- Slide078

By adjusting the threshold (eg, from 0.5 to 1), we have smaller or larger red region (eg, predicted tumor probability > 0.5 to predicted tumor probability =1) depending on how specialists want to penalize the false negative case.



- Slide078 cancer cell ground truth



- Red region shows where predicted tumor probability = 1

Code Walkthrough

```
# 1) use data augmentation to enlarge the unhealthy slide datasets
# from the paper, we defaultly use 2 augmentation methods:
# First, rotate the input patch by 90∘, apply a left-right flip and repeat the rotations.
# Next, we use TensorFlow’s image library to perturb color: brightness with a maximum delta
# of 64/255, saturation with a maximum delta of 0.25, hue with a maximum delta
# of 0.04, and contrast with a maximum delta of 0.75
def color_augmentation(image, brightness = 64/255, saturation = 0.25, hue = 0.04, contrast = 0.75):
  temp = tf.image.random_brightness(image, brightness)
  temp = tf.image.adjust_saturation(temp, 0, saturation)
  temp = tf.image.random_hue(temp, hue)
  result = tf.image.random_contrast(temp, 0, contrast)
  return result

def rotation(image):
  temp = tf.image.rot90(image)
  temp = tf.image.flip_left_right(temp)
  result = tf.image.rot90(temp)
  return result

def cancer_image_processing(image_path):
  img = tf.io.read_file(image_path)
  img = tf.image.decode_jpeg(img, channels=3)
  img = tf.image.resize(img, (299, 299))
  img1 = color_augmentation(img)
  img2 = rotation(img)
  activations1 = tf.keras.applications.inception_v3.preprocess_input(img1)
  activations2 = tf.keras.applications.inception_v3.preprocess_input(img2)
  return activations1, activations2
```

```
[ ] # for train_validation set, read health slide
def read_health_slide(paths):
  X = []
  idx = 0
  while idx < len(paths) - 3:
```

Future Work

- Metrics: add more metrics: eg. precision-recall metrics
- Model: Fine tune the existing Inception V3 model and try ensemble method(includes more models and give majority vote)
- Note that the current test accuracy seems very high, we suspect this happens since majority of test dataset are healthy cells.

Source paper:
Detecting Cancer Metastases on Gigapixel Pathology Images
https://arxiv.org/pdf/1703.02442.pdf

Thank you